Annex to Form PCT/ISA/206
COMMUNICATION RELATING TO THE RESULTS OF THE PARTIAL INTERNATIONAL SEARCH

International Application No PCT/US 01/44285 A18

1. The present communication is an Annex to the invitation to pay additional fees (Form PCT/ISA/206). It shows the results of the international search established on the parts of the international application which relate to the invention first mentioned in claims Nos.:
see 'Invitation to pay additional fees'
2. This communication is not the international search report which will be established according to Article 18 and Rule 43.

3. If the applicant does not pay any additional search fees, the information appearing in this communication will be considered as the result of the international search and will be included as such in the international search report.

4. If the applicant pays additional fees, the international search report will contain both the information appearing in this communication and the results of the international search on other parts of the international application for which such fees will have been paid.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00 37608 A (SLOAN KETTERING INST CANCER) 29 June 2000 (2000-06-29) the whole document	1-26, 29-32
A	THORSON J S ET AL: "Enediyne biosynthesis and self-resistance: A progress report" BIOORGANIC CHEMISTRY, ACADEMIC PRESS INC., NEW YORK, NY, US, vol. 27, no. 2, 1999, pages 172-188, XP002222216 ISSN: 0045-2068 page 178, paragraph 3 -page 184, line 2; figure 7	1-23,25, 29-32
P,X	THORSON ET AL.: "Understanding and exploiting nature's chemical arsenal: the past, present and future of calicheamicin research" CURRENT PHARMACEUTICAL DESIGN, vol. 6, December 2000 (2000-12), pages 1841-1879, XP001146376 page 1860, right-hand column, paragraph 2 -page 1869, right-hand column, paragraph 1	1-26, 29-32

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

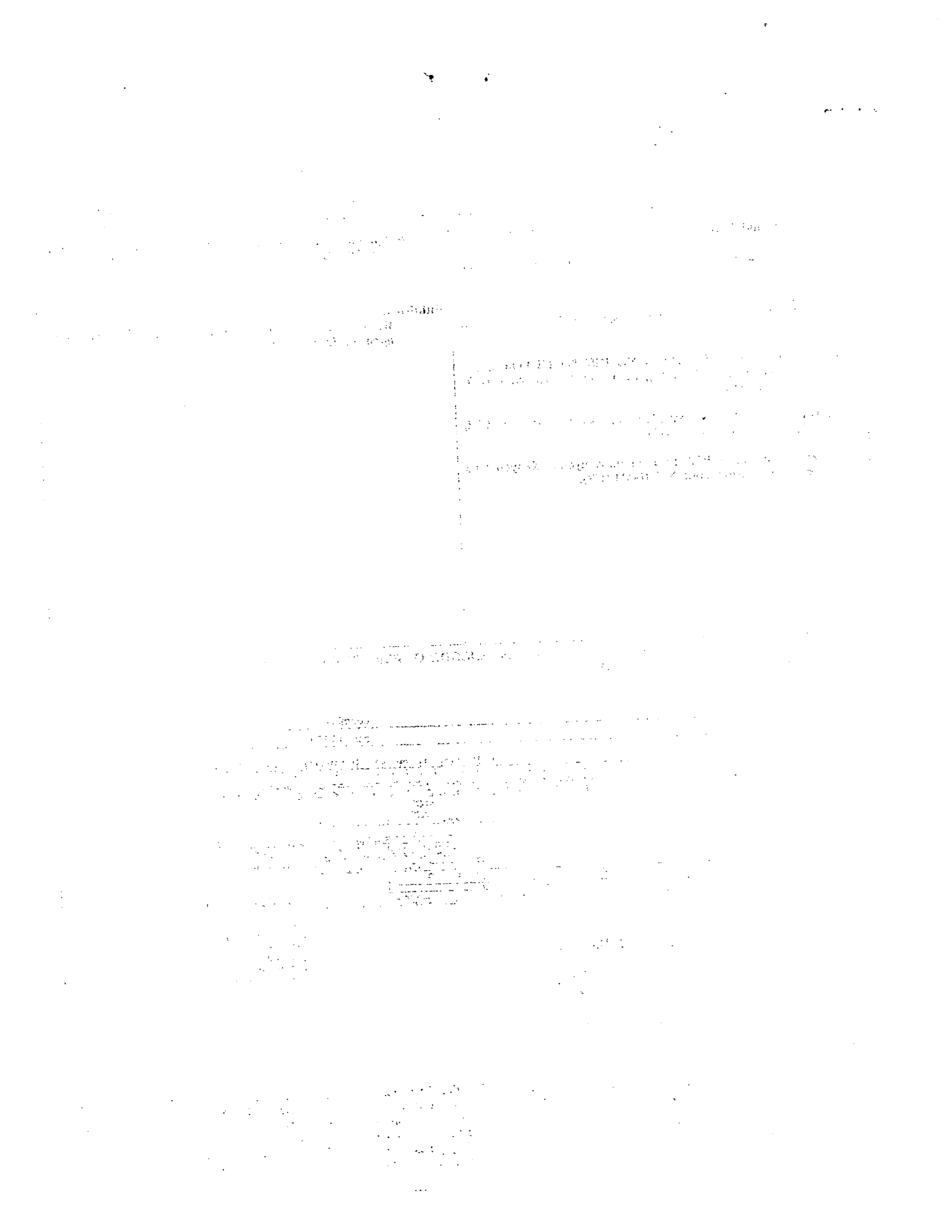
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family



PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N | A2

(11) International Publication Number: **WO 00/37608**

(43) International Publication Date: 29 June 2000 (29.06.00)

(21) International Application Number: PCT/US99/29110

(22) International Filing Date: 7 December 1999 (07.12.99)

(30) Priority Data:
60/111,325 7 December 1998 (07.12.98) US

(71) Applicant: SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH [US/US]; 1275 York Avenue, New York, NY 10021 (US).

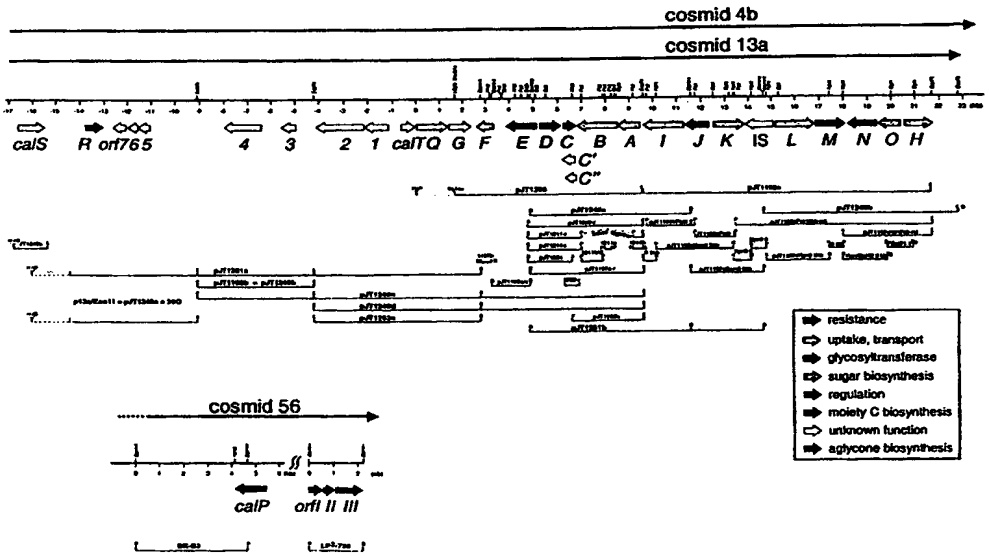
(72) Inventor: THORSON, Jon; 1161 York Avenue, Apartment 3E, New York, NY 10021 (US).

(74) Agents: DELUCIA, Richard, L. et al.; Kenyon & Kenyon, One Broadway, New York, NY 10004 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: *MICROMONOSPORA ECHINOSPORA* GENES ENCODING FOR BIOSYNTHESIS OF CALICHEAMICIN AND SELF-RESISTANCE THERETO



(57) Abstract

An isolated gene cluster of *Micromonospora echinospora* which codes for calicheamicin biosynthesis. The biosynthetic gene cluster contains genes encoding for proteins and enzymes used in the biosynthetic production of calicheamicin, including the aryltetrasaccharide and aglycone. The gene cluster also includes the gene conferring calicheamicin resistance. The invention also provides isolated genes of the biosynthetic cluster and their corresponding proteins. In addition, the invention relates to DNA hybridizing with the calicheamicin gene cluster and the isolated genes of that cluster. Expression vectors containing genes of the biosynthetic gene and their functional variants are also provided. The invention also relates to host cells conjugated with DNA isolated from the *Micromonospora echinospora* spp. *calichensis* genome.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Micromonospora echinospora genes encoding for biosynthesis of calicheamicin and self-resistance thereto

This application claims benefit from provisional application 60/111,325 filed on December 7, 1998, which application is incorporated herein by reference in its entirety.

Field of the Invention

The present invention relates to a biosynthetic gene cluster of *Micromonospora echinospora* spp. *calichensis*. In particular, the calicheamicin biosynthetic gene cluster contains genes encoding for proteins and enzymes used in the biosynthetic pathway and construction of calicheamicin's aryltetrasaccharide and aglycone, and the gene conferring calicheamicin resistance. The present invention also relates to isolated genes of the biosynthetic cluster and their corresponding proteins. In addition, the invention relates to DNA hybridizing with the calicheamicin gene cluster and the isolated genes of that cluster. The invention also relates to expression vectors containing the biosynthetic gene cluster, the individual genes, or functional variants thereof.

Background of the Invention

The enediyne antibiotics, which were discovered in the 1980's, have long been appreciated for their novel molecular architecture, their remarkable biological activity, and their fascinating mode of action. Enediyne antibiotics were originally derived by fermentation of microorganisms, including *Micromonospora, Actinomedura,* and *Streptomyces.* Rothstein, D. M., *Enediyne Antibiotics as Antitumor Agents,* p. 2 (1995).

As a class, the enediyne antibiotics have been referred to as the most potent and highly active antitumor reagents yet discovered. Rothstein, D. M., *Enediyne Antibiotics as Antitumor Agents*, preface (1995).

To date, at least twelve members of this family of antibiotics have been discovered, all of which fall roughly into two categories. The first category of enediynes is classified as chromoprotein enediynes because they possess a novel 9-membered ring chromophore core structure, which also requires a specific associated protein for chromophore stabilization. The second category of enediyne is classified as non-chromoprotein enediynes. These enediynes contain a 10-membered ring, which requires no additional stabilization factors. This enediyne ring structure is often referred to as the "warhead." The warhead induces DNA damage, which is frequently a double-stranded cleavage and appears to be irreparable. This type of DNA damage is usually nonrepairable for the cell and is most often lethal. Because of these remarkable chemical and biological properties, there has been an intense effort by both the pharmaceutical industry and academia to study these substances with the goal of developing new and clinically useful therapeutic anti-tumor agents.

The 9-membered ring chromoprotein enediyne subfamily is comprised of: neocarzinostatin from *Streptomyces carzinostaticus*, (Myers, A.G., et al., *J. Am. Chem. Soc., 110,* 7212-7214 (1988)); kedarcidin from *Actinomycete* L585-6, (Leet, J.E., et al., *J. Am. Chem. Soc., 114,* 7946-7948 (1992)), N1999A2 from *Streptomyces globisporus*, (Yoshida, K., et al. *Tetrahedron Lett., 34,* 2637-2640 (1993)), maduropeptin from *Actinomadura madurea*, (Schroeder, D.R., et al., *J. Am. Chem. Soc.*, 116, 9351-9352 (1994)); N1999A2 from *Streptomyces sp.* AJ9493, (Schroeder, D.R., et al., *J. Am. Chem.*

Soc., 116, 9351-9352 (1994)): actinoxanthin from *Actinomyces globisporus*, (Khokhlov, A.S., et al., *J. Antibiot., XXII,* 541-544 (1969)); largomycin from *Streptomyces pluricolorescens*, (Yamaguchi, T., et al., *J. Antibiot., XXIII,* 369-372 (1970)); auromomycin from *Streptomyces macromomyceticus,* (Yamashita, T., et al., *J. Antibiot., XXXII,* 330-339 (1979)), and sporamycin from *Streptosporangium pseudovulgare*, (Komiyama, K, et al., *J. Antibiot.,* XXX, 202-208 (1977)) all of which are believed to possess a novel bicylo[7.3.0.]dodecadiynene chromophore core structure essential for biological activity. In addition, with the exception of N1999A2, a required apoprotein acts as a stabilizer and specific carrier for the unstable chromophore, and for its transport and interaction with target DNA.

The non-chromophore enediyne subfamily is comprised of calicheamicin from *Micromonospora echinospora* spp. *calichensis*; namenamicin from *Polysyncraton lithostrotum;* esperamicin from *Actinomadura verrucosospora;* and dynemicin from *Micromonospora chersina.*

Enediyne antibiotics have potential as anticancer agents because of their ability to cleave DNA, however, many of these compounds are too toxic to be used currently in clinical studies. Today, only calicheamicin is known to be currently used in clinical trials and it has provided promising results as an anticancer agent. The enediynes potentially have utility as anti-infective agents, provided that toxicity can be managed.

The toxicity of the enediyne compounds, including calicheamicin, centers on the problem of directing the compound to the cleave only the DNA of interest, such as tumor cell DNA, and not the DNA of the host. Due to calicheamicin's powerful ability to cleave DNA, scientists have investigated the mechanism by which calicheamicin-producing

organism protects itself against the DNA-cleaving activity of the molecule. Rothstein, D. M., *Enediyne Antibiotics as Antitumor Agents*, p. 77 (1995). Prior to this invention, knowledge of genes encoding for non-chromoprotein enediyne self resistance was completely lacking. Insight into how *Micromonospora* self resistance gene and gene products act to control the toxic effects of calicheamicin offers new avenues of clinical research. For example, knowledge of the mechanisms underlying calicheamicin resistance could provide the means necessary to use higher doses of calicheamicin while simultaneously inhibiting the toxic effects of the drug on non-cancer cells. Additionally, understanding the mechanism behind calicheamicin's self-resistance may aid in the understanding of self-resistance in other enediyne antibiotics, thereby potentially making useful those enediynes once thought to be too toxic to be viably used as therapeutic agents. The calicheamicin self-resistance mechanisms elucidated utilizing the present invention provide gene therapy approaches, for example, via introduction of enediynes resistance genes into bone marrow cells, thereby increasing resistance and allowing tolerance to chemotherapeutic doses of calicheamicin. Banerjee, D., et al., *Stem Cells*, 12, 378-385 (1994). Thus, understanding calicheamicin self-resistance will significantly aid continuing clinical studies involving calicheamicin and the enediynes. The present invention addresses this need as it provides for the isolation and characterization of a resistance gene and its associated protein for any nonchromoprotein enediynes.

Calicheamicin has two distinct structural regions: the aryltetrasaccharide and the aglycone (also known as the warhead). The aryltetrasaccharide displays a highly unusual series of glycosidic, thioester, and hydroxylamine linkages and serves to deliver the drug to specific tracts (5'-TCCT-3' and 5'-TTTT-3') within the minor groove of DNA. The

aglycone of calicheamicin consists of a highly functionalized bicyclo[7.3.1]tridecadiynene core structure with an allylic trisulfide serving as the triggering mechanism. McGahren, W.J.,et al., *Enediyne Antibiotics as Antitumor Agents*, pp. 75-86 (1995). Once the aryltetrasaccharide is firmly docked, aromatization of the bicyclo[7.3.1]tridecadiynene core structure, via a 1,4-dehydrobenzene-diradical, results in the site specific oxidative double strand scission of the targeted DNA. Zein, N.,et al., *Science*, 240, 1198-1201 (1988). The aglycone undergoes a reaction that yields carbon-centered diradicals, which are responsible for DNA cleavage. This activity has sparked considerable interest in the pharmaceutical industry culminating in the recent success of calicheamicin-antibody conjugates (CMA-676) to treat acute myelogenous leukemia (AML) in phase III trials. Additionally, similar strategies have been used in phase I trials to treat breast cancer. A massive program to examine calicheamicin conjugated to alternative delivery systems has also recently been undertaken. Hamann, P.R., et al., *87th Annual Meeting of the American Association of Cancer Research*, Washington, D.C., pp. 471 (1996); Hinman, L.M., et al., *Cancer Res.*, *53*, 3336 (1993); Hinman, L. M., et al., *Enediyne Antibiotics as Antitumor Agents*, pp. 87- 105 (1995); Sievers, E.L., et al., *Blood*, *93*, 3678-3684 (1999); Siegel, M.M., et al., *Anal. Chem.*, *69*, 2716-2726 (1997); Ellestad, G. personal communication.

The biological activity and molecular architecture of calicheamicin has also prompted a search for the potentially useful analogs. Of the numerous laboratories producing synthetic analogs, one group has produced a novel calicheamicin θ^{I}_{1} shown to effectively suppress growth and dissemination of liver metastases in a syngeneic model of murine neuroblastoma. Lode, H. N., et al., *Cancer Res.*, *58*, 2925-2928 (1998); Wrasidlo,

W., et al., *Acta Oncologica, 34*, 157-164 (1995). In addition to synthesizing calicheamicin analogs, random mutagenesis of *M. echinospora* and screening for mutant strains with improved biosynthetic potential has also been pursued. Rothstein, D. M., *Enediyne Antibiotics as Antitumor Agents*, pp. 107-126 (1995).

The first total synthesis of calicheamicin was reported by Nicolaou and coworkers in 1992. Synthesizing this complex antibiotic, though, presents many disadvantages. For example, Nacelle's procedure only provides approximately a 0.007% yield and requires 47 steps. Halcomb, R.L., Enediyne Antibiotics as Antitumor Agents, pp. 383-439 (1995). Thus, the total synthesis of calicheamicin remains secondary to the isolation of calicheamicin from large fermentations of *M. echinospora*. Therefore, methods to produce mass amounts of calicheamicin and potentially useful variants are still needed. Fantini, A., et al., *Enediyne Antibiotics as Antitumor Agents*, pp. 29-48 (1995). Transforming calicheamicin DNA into producing strains of bacteria, *E. coli* for example, would address this need. Currently there are no cloned *M. echinospora* genes and only a set of limited studies upon putative *M. echinospora* promoters are available. Lin, L.S., et al., *J. Gen. Microbiol., 138,1881*-1885 (1992); Lin, L.S., et al., *J. Bacteriol.*, 174, 3111-3117 (1992); Baum, E.Z., et al., *J. Bacteriol.*, 171, 6503-6510 (1989); Baum, E.Z., et al., *J. Bacteriol.*, 170, 71-77 (1988).

Having calicheamicin DNA opens the door for genetic analysis of calicheamicin biosynthesis, as such analysis requires the ability to obtain large qualities of calicheamicin DNA. For example, one can study calicheamicin biosynthesis by mutagenesis of *M. echinospora*, including the isolation and characterization of mutants blocked in calicheamicin biosynthesis and the subsequent analysis of their defective or partial

calicheamicin products. Additionally, particular enzyme could be overexpressed or underexpressed after subcloning its gene into a host such as *E. coli*, and the results of such overexpression studied to reveal the enzyme's function. Furthermore, the cloning of biosynthetic genes can ultimately result in increased yields of the gene product by cloning and expressing the biosynthetic gene encoding the rate-limiting enzyme back into the producing organism. It may also be possible to generate novel products by cloning biosynthetic genes into strains that make related compounds. Such genes could endow the host organism with the ability to carry out new reactions on the enediyne nucleus, and thus produce novel drugs.

Calicheamicin's molecular architecture in conjunction with its useful biological activity and potential therapeutic value brand calicheamicin an target for the study of natural product biosynthesis. While the radical-based mechanism of oxidative DNA cleavage by calicheamicin (i.e. aromatization of the bicyclo[7.3.1]tridecadiynene core structure, via a 1,4-dehydrobenzene-diradical, resulting in the site specific oxidative double strand DNA cleavage) is well understood, it was unknown, prior to this invention, how *Micromonospora* constructs calicheamicin. As a result, there is a need to discover and understand calicheamicin biosynthesis. Prior to this discovery, knowledge of genes encoding for nonchromoprotein enediyne biosynthesis was completely lacking. Thus, this invention relates to the first identification, isolation, and cloning of a nonchromoprotein enediyne biosynthetic gene cluster and mapping and nucleotide sequence analysis of the genes within the cluster. The invention provides the entire calicheamicin-biosynthetic cluster and biochemical studies of aryltetrasaccharide biosynthesis. Furthermore, the calicheamicin self-resistance gene and protein have been

isolated as have the genes and resulting enzymes for steps within the calicheamicin cascade. The invention also provides for construction of enediyne overproducing strains for rational biosynthetic modification of bioactive secondary metabolites, for new drug leads, and for an enediyne combinatorial biosynthesis program.

The present invention thus also relates to a biosynthetic modification of bioactive secondary metabolites through enediyne combinatorial biosynthesis. As most pharmaceutical drug leads are inspired by naturally occurring compounds, and given the challenge posed in synthesizing these metabolites, genetic manipulation of the sugar appendage on the metabolites offers avenues for creating potential new drugs. Thus the emerging field of combinatorial biosynthesis has become a rich new source for modified non-natural sugar scaffolds. Marsden, A., et al., *Science* 1998, 279, 199-201. Problems inherent with the genetic manipulation of the sugar appendage relate to the fact that naturally occurring bioactive secondary metabolites possess unusual carbohydrate ligands, which serve as molecular recognition elements critical for biological activity. *Macrolide Antibiotics: Chemistry, Biology and Practice, 1984.* Without these essential sugar attachments, the biological activities of most clinically important secondary metabolites are either completely abolished or dramatically decreased. Currently, techniques for the genetic manipulation of the sugar appendage for a given metabolite rely mainly on the alteration and/or deletion of a small subset of genes required to construct and attach each desired sugar moiety. Thus there is a need to develop alternate strategies to construct and attach non-naturally occurring sugars. The present invention addresses this need. The present invention utilizes the fact that glycosyltransferases, which are responsible for the final glycosylation of certain secondary metabolites, show a

high degree of promiscuity toward the nucleotide sugar donor. Zhao, L., et al., *J. Am. Chem. Soc.* 1988, 120, 12159-12160. This unselectivity of the glycosyltransferases has the potential for allowing modification of the crucial glycosylation pattern of natural, or non-natural, secondary metabolite scaffolds in a combinatorial fashion. The present invention discloses a method using the recruitment and collaborative action of sugar genes from a variety of biosynthetic pathways to construct composite gene clusters, which make and attach non-natural sugars.

Summary of the Invention

The present invention provides an isolated nucleic acid molecule from *Micromonospora echinospora* encoding for a gene from a nonchromoprotein enediyne biosynthetic gene cluster, the protein coding region of the gene or a biologically active fragment of the gene. In particular, the present invention provides an isolated nucleic acid molecule, gene, or gene cluster from *Micromonospora echinospora* spp. *calichensis* that is involved in the biosynthesis of calicheamicin. In another embodiment, the present invention also relates to nucleic acids capable of hybridizing with a nucleic acid molecule from *Micromonospora echinospora* spp. *calichensis* coding for one or more genes from a nonchromoprotein enediyne biosynthetic gene cluster. In a further embodiment the invention provides an expression vector comprising an isolated nucleic acid molecule from a nonchromoprotein enediyne biosynthetic gene cluster from *Micromonospora echinospora*. In yet a further embodiment the invention provides a cosmid comprising an isolated nucleic acid molecule from *Micromonospora echinospora* comprising a nucleic acid sequence encoding for a nonchromoprotein enediyne biosynthetic gene cluster.

In preferred embodiments, the invention provides the isolated nucleic acid molecules of SEQ ID Nos. 1, 3, and 5.

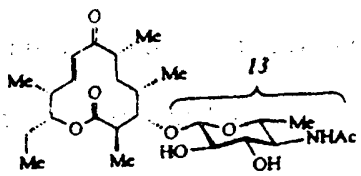
In an additional embodiment, the present invention provides a host cell transformed with an isolated nucleic acid molecule from a nonchromoprotein enediyne biosynthetic gene cluster from *Micromonospora echinospora*. Host cells can optionally be of bacterial, yeast, fungal, insect, plant or mammalian origin and can be transformed according to standard methods. In a preferred embodiment, the host cell is the bacterium *E. coli* or *Streptomyces*. In a further embodiment, the invention is directed to a transformed host cell with an expression vector encoding gene *cal*C, or a functional derivative thereof, operably linked to regulatory sequences that enable expression of *cal*C. In a yet further embodiment, the invention provides a transformed host cell with an expression vector encoding the gene *cal*H, or a functional derivative thereof, operably linked to regulatory sequences that enable expression of *cal*H. Likewise, the invention provides a transformed host cell with an expression vector encoding the gene *cal*G, or a functional derivative thereof, operably linked to regulatory sequences that enable expression of *cal*G.

The invention further provides a method of expressing a protein by culturing a host cell transformed with an expression vector comprising an isolated nucleic acid molecule from *Micromonospora echinospora* encoding for a gene from a nonchromoprotein enediyne biosynthetic gene cluster, and incubating the host cell for a time and under conditions allowing for protein expression. In another embodiment the invention provides a method of purifying calicheamicin using affinity chromatography. A sample containing calicheamicin is contacted with an affinity matrix having the protein

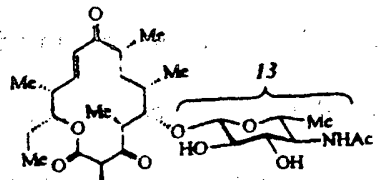
CalC bound thereto, for a time and under conditions allowing calicheamicin to bind to the matrix, eluting calicheamicin from the matrix, and recovering calicheamicin.

In a further embodiment the present invention provides polypeptides having the amino acid sequences of SEQ ID Nos. 2, 4, and 6.

In yet a further the invention provides the production of the following two new macrolides:



11



12

The invention further provides a method of conferring calicheamicin resistance to a subject comprising obtaining cells from the subject, transforming the cells with the calicheamicin self-resistance gene, and returning the cells to the subject. Alternatively, the calicheamicin self-resistance gene can be targeted and delivered to the desired host cells through known gene therapy delivery systems.

Brief Description of the Figures

Figure 1 depicts the summary of the cosmid clones isolated from *M. echinospora* genomic library. This figure illustrates the results of the screening of the genomic library for clones carrying the calicheamicin biosynthetic cluster.

Figure 2 shows a restriction map of a portion of cosmid clones 4b, 13a, and 56 and the corresponding location of *cal* genes from *M. echinospora*.

Figure 3 is a table of the open reading frames ("orfs") in the calicheamicin biosynthetic cluster. This table lists the polypeptides that the genes encode for as well as their proposed or actual determined function in the biosynthetic pathway.

Figure 4 is a graph of the UV-visible absorption spectra of purified mbp-CalC. The purified mpb-CalC was analyzed in the following solution: 52 μM mpb-CalC; 10 mM Tris-HCl, pH 7.5). The inset shows the results of low temperature (4.3 K) the X-band EPR analysis of CalC. 250 μM mpb-CalC containing 0.5 mol Fe per mol CalC was analyzed in 10 mM Tris-HCl, pH 7.5. The spectrometer settings were as follows: field set = 2050 G; scan range = 4,000G; time constant = 82 s; modulation amplitude =16 G; microwave power = 31 μW; frequency = 9.71 Ghz; gain = 1000; determined spin quantitation = 90 ± 10 μM Fe.

Figure 4(b) provides the results of the mbp-CalC *in vitro* assay.

Figure 5 depicts the postulated routes for the biosynthesis of required nucleotide sugars. The enzymes are depicted as follows: E_{deox} = deoxygenase; E_{am} = aminotransferase; E_{ep} = epimerase; E_{me} = methyltransferase; E_{od} = 4,6-dehydratase; E_{ox} = oxidase; E_p = nucleotidyltransferase; E_{red} = reductase; E_{sh} = sulfhydrytransferase.

Figure 6 illustrates a schematic representation of the *in vivo* production of pikromycin/methymycin-calicheamicin hybrid metabolites.

Figure 7 depicts the *Streptomyces venezuela* methymycin/pikromycin gene cluster. Eight open reading frames (*des*I-*des*VIII) in this cluster have been assigned as genes involved in desoamine biosynthesis. This figure also depicts the hybrid pathway toward new methymycin/pikromycin derivatives (11 and 12) produced after heterologous expression of the *cal*H gene of calicheamicin in a *S. venezuela* mutant.

Figure 8 illustrates calicheamicin's (6) four unique sugars which are crucial to tight DNA binding. Sugar (9) is derived from 4-amino-4,6-dideoxyglucose (8) and is part of the restricted N-O connection between sugars A and B. Compound 8 is derived from the corresponding 4-ketosugar (7) via a transamination reaction. The gene *cal*H encodes the desired C-4 aminotransferase allowing conversion of compound (7) to compound (8).

Detailed Description of the Invention

The present invention is directed to the isolation and characterization of the calicheamicin biosynthetic cluster. This cluster encodes the genes that encode the proteins and enzymes that are involved in the deoxysugar synthesis (the aryltetrasaccharide), polyketide biosynthesis (the aglycone) of calicheamicin synthesis, and calicheamicin resistance. Twenty-one structural genes have been identified that encode for the aryltetrasaccharide sugar ligands (~20 kb); approximately eight modules (~40 kb) are required for the 15-carbon aglycone. Four proteins involved in transport and uptake, one protein conferring resistance, and one regulatory protein have been identified.

The calicheamicin biosynthetic gene cluster comprises the following genes: *cal*A, *cal*B, *cal*C, *cal*D, *cal*E, *cal*F, *cal*G, *cal*H, *cal*I, *cal*J, *cal*K, *cal*L, *cal*M, *cal*N, *cal*O, *cal*P, *cal*Q, *cal*R, *cal*S, *cal*T, *orf*1, *orf*2, *orf*3, *orf*4, *orf*5, *orf*6, *orf*7, and an IS-element gene. The above listed genes encode for the following polypeptides: CalA (328 amino acids), CalB (561 amino acids), CalC (181 amino acids), CalD (263 amino acids), CalE (420 amino acids), CalF (245 amino acids), CalG (990 amino acids), CalH (338 amino acids), CalI (568 amino acids), CalJ (332 amino acids), CalK (440 amino acids), Cal L (562 amino acids), Cal M (416 amino acids), CalN (398 amino acids), CalO (331 amino acids),

Cal P (approximately 179 amino acids), CalQ (453 amino acids), CalR (265 amino acids), CalS (1113 amino acids), CalT (280 amino acids), Orf1(322 amino acids), Orf2 (654 amino acids), Orf3 (209 amino acids), Orf4 (521 amino acids), Orf5 (175 amino acids), Orf6 (139 amino acids), Orf7 (187 amino acids), and IS-element (402 amino acids).

In elucidating the calicheamicin biosynthetic gene cluster, the inventors began with a genomic library containing the genome of *Micromonospora echinospora* spp. *calichensis*. The cosmid library was generated by isolating chromosomal DNA of *Micromonospora echinospora* spp. *calichensis*, fragmenting that chromosomal DNA, inserting the DNA into a cosmid vector and generating a cosmid library according to methods well known in the art. This procedure can be performed in any species of *Micromonospora*.

Based upon prior enediyne metabolic labeling studies it was postulated that the calicheamicin aglycone would be polyketide derived. Polyketide metabolites encompass a vast variety of structural diversities yet share a common mechanism of biosynthesis. Hutchinson, C.R., et al., *Chem. Rev., 97*, 2525-2535 (1997); Strohl, W.R., et al, *Biotechnology of Antibiotics* pp. 577-657; Fujii, I., et al., *Chem. Rev., 97*, 2511-2523 (1997); Hopwood, D.A., et al., *Chem Rev., 97*, 2465-2497 (1997); Hopwood, D.A., et al., *Ann. Rev. Genet., 24*, 37-66 (1990); Staunton, J., et al., *Chemical Reviews, 97*, 2611-2629 (1997). Most important, polyketide synthase ("PKS") genes display a high degree of sequence homology (from pathway to pathway and organism to organism) and are often clustered with genes encoding self resistance and deoxysugar ligand biosynthesis. Hopwood, D.A., et al., *Chem. Rev., 97*, 2465-2497 (1997); Hopwood, D.A., et al., *Ann. Rev. Genet., 24*, 37-66 (1990); Staunton, J., et al., *Chem. Rev., 97*, 2611-2629 (1997).

Degenerate primers based upon conserved regions within PKS genes were used in Southern hybridizations to identify clones from the *M. echinospora* genomic library that carried putative PKS genes. The Southern hybridizations were performed by methods known in the art. Southern hybridization of the genomic *M. echinospora* cosmid library with a DNA probe designed to target type I PKS genes (KS^I). (Kakavas, S.J., et al., *J. Bacteriol., 179,* 7515-7522 (1997)), unveiled five positive clones, which were designated clones 4b, 10a, 13a, 56, and 60. See Figure 1. The same five clones were also identified upon rescreening the genomic library with type II DNA probe (actI). See Figure 1. Although this preliminary analysis clearly demonstrated the presence of Micromonospora PKS gene homologues, a secondary screen was performed as PKS hybridization analyses are often plagued by false hybridization to gene clusters that encode spore pigment biosynthesis.

The second screening was based on the assumption that calicheamicin's biosynthetic cluster would also contain genes encoding for deoxysugar ligand synthesis. Further, it was postulated that all hexopyranosyl ligands of calicheamicin diverged from the common intermediate 4-keto-6-deoxy TDP-D-glucose (30), Figure 5, as macromolecule-sugar synthesis in many organisms began with a similar common intermediate. Thus, it was believed that the cluster encoding for calicheamicin biosynthesis should, in addition to carrying a PKS-encoding region, would carry both a common glucose-l-phosphate nucleotidyltransferase and a NDP-α-D-glucose 4,6-dehydratase gene, encoding the putative enzymes E_{pl}, and E_{od}, respectively. See figure 5. These enzymes are necessary to convert a sugar (12)(figure 5) to the hypothesized common intermediate, 4-keto-6-deoxy TDP-D-glucose (30). Analogs to 4,6-dehydratases

have been previously characterized from *E. coli, Salmonella.* and *Streptomyces.* Additionally, a nucleotide transferase from *Salmonella* has been characterized as an alpha-D-glucose-1-phosphate thymidylyltransferase. The secondary screen was performed using a probe based upon the postulation that the M. echinospora's calicheamicin synthesis would begin from a similar precursor found in *E. coli, Streptomyces* and *Salmonella*, and that this precursor required a dehydratase to convert it into the common intermediate, 4-keto-6-deoxy TDP-D-glucose (30). In particular, a DNA probe (designated E_{od}^{1}) was designed from the conserved NAD^{+}-binding site of bacterial NDP-α-D-glucose 4,6-dehydratases. He, X., et al., *Biochem., 35,* 4721-4731 (1996). Southern hybridization of the genomic *M. echinospora* cosmid library with the E_{od}^{1} probe revealed cross-hybridization with clones 4b, 10a, 13a, 56, and 60. Two additional clones, designated 58 and 66, were also identified in this screen. See Figure 1. This secondary hybridization indicated the clustering of genes encoding both polyketide and deoxysugar biosynthesis.

For final corroboration, since secondary metabolite biosynthesis is typically clustered with resistance genes in actinomycetes, all hybridization-positive clones were tested for their ability to grow in the presence of varying concentrations of calicheamicin. In this final screen, six of the seven hybridizing clones displayed differing levels of resistance to calicheamicin (4b≈10a≈13a≥56≥66>60)(See Figure 1) while clone 58 lacked the ability to grow in the presence of calicheamicin. In addition, these resistance screens revealed that clones 4b, 10a, 13a conferred much higher levels of resistance to calicheamicin than the other clones. Upon rescreening the genomic library for calicheamicin-resistant clones, three additional clones (3a, 4a, and 16a) were found to

confer similar levels of resistance. Cumulatively, the results demonstrated that clones 4b, 10a, 13a, 56, and 60 carried PKS I and II homologues and deoxy sugar biosynthetic genes, as well as encoded the gene responsible for conferring calicheamicin-self resistance.

The clones positive for PKS I and II and deoxy sugar biosynthesis homology and calicheamicin resistance were used to map the biosynthetic cluster. Southern hybridization established similarity between clones 3a, 4a, 4b, 10a, 13a, 16a and 56. In addition, nucleotide sequence overlaps were found between clones 4b, 13a, and 56. See Figure 1. Restriction mapping and Southern hybridization of these clones indicated that the positive cosmid clones corresponded to a continuous region of the *M. echinospora* chromosome spanning > 100 kb. The present invention thus provides for cosmids having a nucleic acid molecule from Micromonospora echinospora encoding for a nonchromoprotein enediyne biosynthetic cluster.

After isolating the biosynthetic gene cluster and elucidating the sequence, open reading frames ("orfs") were assigned. Tentative gene assignments were derived from amino acid sequence similarity of translated orfs to gene products of known function via direct BLAST (Basic Local Alignment Search Tool) database searches on the amino acid level. Karlin, et al., *Proceed Natl. Acad. Sci., U.S.A., 87,* 2264-2268 (1990); Karlin, et al., *Proceed Natl. Acad. Sci., U.S.A., 90,* 5873-5877 (1993); Altchul, *Nature Genet., 6,* 119-129 (1994). The gene cluster organization is provided in figure 1.

Based on BLAST analysis tentative gene assignments were made. It was deducted that genes participating in the construction of the aryltetrasaccharide include: a) genes encoding nucleotide sugar biosynthesis *(calG,* H, K, O, Q, and S); b) genes

encoding for aryltetrasaccharide assembly (*cal*E and N); and c) genes encoding for "tailoring" reactions (*cal*D, F, and J).

One aspect of the invention relates to transformation of a host cell with *M. echinospora* DNA. This method provides a reproducible transformation efficiency of ~10^3 kanamycin resistant transformants/μg DNA using a pKC1139-based vector. The invention further provides that the host cell can be but is not limited to bacteria, yeast, fungus, insect, plant or mammalian. Transformations of bacteria, yeast, fungus, insect, plant or mammalian cells are performed by methods known in the art.

The present invention also provides the isolation and characterization of the gene encoding for calicheamicin resistance. One aspect of the invention relates to an isolated DNA strand having the gene *cal*C and having the DNA sequence SEQ. ID No.: 1. The present invention also relates to an isolated protein CalC, having the amino acid sequence, SEQ ID. NO. 2. The invention further provides for *cal*C gene fragments coding for a bioactive CalC. The polypeptide, CalC, confers calicheamicin resistance and has 181 amino acids. The invention also provides for CalC fragments conferring calicheamicin resistance.

The *cal*C locus was isolated by identifying calicheamicin genomic cosmid clones that were able to grow on luria bertani ("LB") agar plates containing ampicillin and calicheamicin. The DNA of the positive clones (clones that grew on the plates containing calicheamicin) was isolated and subsequent restriction mapping localized the desired phenotype (calicheamicin resistance). The DNA was then sequenced and the open reading frames analyzed to ascertain the orf encoding for the desired phenotype. *In vitro* studies were also performed and confirmed the ability of CalC to inhibit DNA cleavage.

DNA containing *calC* was cloned into an inducible vector, using known methods, resulting in overexpression of *calC*. The polypeptide product (CalC) was then isolated and purified to homogeneity. Analysis of the purified CalC revealed that it is a non-heme iron metalloprotein that functions via inhibition of calicheamicin-induced DNA cleavage *in vitro*. Another aspect of the invention is an expression vector containing *calC* or a fragment of *calC* encoding for a bioactive molecule. There is also provided a transformed host cell, preferably bacteria, more preferably *E. coli* containing *calC* or a fragment of *calC* encoding for a bioactive molecule.

The present invention provides for the transformation of human cells with the *calC* gene. This allows bone marrow cells, for example, to be removed from a patient being treated with calicheamicin, and to transform these cells with *calC*, and return the transformed cells to the patient. This allows the patient to tolerate treatment with calicheamicin or allows the patient to receive higher doses of calicheamicin as the returned human-*calC*-transformed cells have calicheamicin resistance. The transformation is performed by methods known in the art. The embodiment of the invention would be applicable to many diseases being treated with calicheamicin.

Another aspect of the invention relates to an isolated DNA strand containing the *calH* gene having the DNA sequence SEQ ID. No: 3. The invention also relates to the polypeptide CalH, having amino acid sequence SEQ ID. No. 4. The invention further provides for *calH* gene fragments coding for a bioactive CalH. CalH is involved in the formation of the aryltetrasaccharide 4,6-dideoxy-4-hydroxylamino-D-glucose moiety. CalH catalyzes the conversion of intermediate (30) to intermediate (39) (figure 5). CalH is a TDP-6-deoxy-D-glycerol-L-threo-4-hexulose 4-transaminase, which catalyzes a

pyridoxal phosphate ("PLP")-dependent transamination from glutamate to provide 4-amino-6-deoxy TDP-D glucose (intermediate 39)(figure 5). The invention also provides for CalH fragments that retain bioactivity. There is also provided an expression vector containing the *cal*H gene or fragments of the *cal*H gene that encode for a bioactive polypeptide. CalH were overexpressed as a $(histidine)_{10}$-fusion protein and subsequently purified by nickel affinity chromatography.

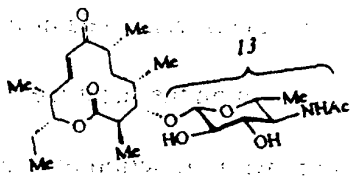
According to BLAST analysis, *cal*H closely resembled perosamine synthase, an enzyme which converts compound 30 to compound 39 (See figure 5) *en route* to the biosynthesis of TDP-perosamine (TDP-4,6-dideoxy-4-amino-D-mannose) in *E. coli.* Wang, L., et al., *Infect. Immunol.*, *66*, 3545-3551 (1998). Thus CalH was believed to be a 4-ketohexose aminotransferase. To confirm the tentative BLAST assigned function, a combinatorial biosynthesis was performed. Specifically the *cal*H gene from calicheamicin was incorporated into a mutant strain of *Streptomyces venezuela.* The 4-dehydrase gene (*des*1) in the methymycin/pikromycin pathway was deleted in this mutant strain. A promoter sequence from the *S. venezuela* methymycin/pikromycin cluster was incorporated in the expression vector to drive the expression of foreign genes (the *cal*H of calicheamicin) in *S. venezuela.* In wild type *S. venezuela* methymycin/pikromycin pathway is known to produce methymycin, neomethymycin, pikromycin, and narbomycin. See figure 6. Deletion of the *des*1 gene in the mutant strain led to the accumulation of the CalH substrate, TDP-4-keto-6-deoxyglucose (compound 30, figure 6). The constructed expression vector with the S. venezuela promoter expressed the *cal*H gene to make the CalH protein. CalH acted on the substrate, 30, to produce compound 39 (figure 6). Compound 39 in turn, with the action of *S. venezuela's* DesVII (a glycosyltransferase)

produced two methymycin/pikromycin-calicheamicin hybrid compounds. See Figure 6, compounds 40 and 41. These hybrid compounds carry the 4-aminohexose ligand of calicheamicin. This work provides indisputable support for the *calH* gene assignment as encoding the TDP-6-deoxy -D-glycero-L-threo-4-hexulose 4-aminotransferase of the calicheamicin pathway. The CalH acted on the TDP-4-keto-deoxyglucose substrate (compound 30) to produce compound 39. (Figure 5).

In addition, these results reinforce the indiscriminate nature of the corresponding glycosyltransferase (DesVII) as it reveals that the glycosyltransferase (DesVII) of the *S. venezuela* pathway can recognize alternative sugar substrates whose structures are considerably different from the original amino sugar substrate, TDP-D-desoamine. The results also clearly demonstrate the ability to engineer secondary metabolite glycosylation through a rational selection of gene combinations. The successful expression of the CalH protein in *S. venezuela* by the newly constructed expression vector highlights the potential of using this system to express other foreign genes in this strain.

Thus, one aspect of the present invention further relates to the construction of a composite gene cluster having the ability to make and attach non-natural sugars. The invention further provides an expression vector having a calicheamicin gene operably linked to regulatory sequences to control expression of the calicheamicin protein and preferably the regulatory sequence is a *Streptomyces* promoter. The present invention also relates to two newly synthesized sugars, compound (11) and compound (12)(figure 7).

Compound 11 has the formula:

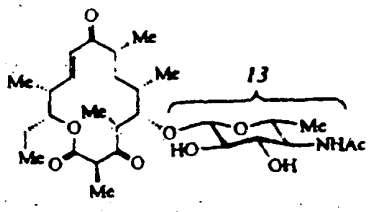


11

The spectral data of compound 11 was as follows:

^{1}H NMR (500 MHz $CDCl_3$, J in hertz) δ 6.75 (1H, dd, J = 16.0, 5.5, 9-H) 6.44 (1H, dd, J = 16.0, 1.2, 8-H), 5.34 (1H, d, j = 8.0, N-H), 4.96 (1H, m, 11-H), 4.27 (1H, d, J=7.5, 1-H), 3.66 (1H, dd, J = 9.5, 8.0, 4'-H), 3.60 (1H, d, J = 10.5, 3-H), 3.50 (1H, 1, J - 9.5, 3'H), 3.[d] (1H, m, 5'-H), 3.4 (1H, m, 2'-H), 2.84 (1H, dq, J = 10.5, 7.5, 2-H), 2.64 (1H, m, 10-H), 2.53 (1H, m, 6-H), 2.06 (3H, s, Me-C=0), 1.7 (1H, m, 12-H), 1.66 (1H, m, 5-H), 1.56 (1H, m, 12-H), 1.4 (1H, M, 5-H), 1.36 (3H, d, J=7.5, 2-Me), 1.25 (311, D, J = 6.5, 5'-Me), 1.24 (1H, m, 4-H), 1.21 (3H, d, J=7.5, 6 Me), 1.10 (3H, d, J=6.5, 10-Me), 0.99 (3H, d, J=6.0, 4-Me), 0.91 (3H, t, J =7.2, 12-Me); ^{13}C NMR (125 MHz, $CDCl_3$) δ 205.3 (C-7), 175.1 (C-1), 171.9 (Me-C-O), 147.1 (C-9), 126.1 (C-8), 103.0 (C-1'), 85.8 (C-3), 75.8 (C-5'), 75.8 (C-3'), 74.1 (C-11) 70.8 (C-2'), 57.6 (C-4'), 45.3 (C-6), 44.0 (C-2), 38.1 (C-10), 34.2 (C-5), 33.6 (C-4), 25.4 (C-12), 23.7 (Me-C-O), 18.1 (C-6'), 17.9 (6 Me), 17.6 (4-Me), 16.4 (2-Me), 10.5 (12-Me), 9.8 (10 Me). High-resolution FAB-MS calculated for $C_{25}H_{42}NO_8$ (M + H$^+$) 484.2910, found 484.2303.

Compound 12 has the formula:



12

The spectral data of compound 12 was as follows:

^{1}H NMR (500 MHz, $CDCl_3$, J in hertz) δ 6.69 (1H, dd, J = 16.0, 6.0, 11-H), 6.09 (1H, dd, J = 16.0, 1.5, 10-H), 5.35 (1H, d, J = 8.5, N-H), 4.96 (1H, m, 13-H), 4.36 (1H, d, J = 7.5, 1'H), 4.19 (1H, m, 5-H), 3.83 (1H-q, J=6.5, 2-H), 3.68 (1H, dt, J=10.0, 8.5, 4'H), 3.52 (1H, t, J = 8.5, 3-'H), 3.50 (1H, m, 5-H), 3.42 (1H, t, J = 7.5, 2'-H), 2.92 (1H, dq, J = 7.0, 5.0, 4-H), 2.81 (1H, m, 8-H), 2.73 (1H, t, J=7.5, 2'-H), 2.06 (3H, a, Me-C-O), 1.8 (1H, m, 6-H), 1.6 (1H, m, 14-H), 1.55 (1H, m, 7-H), 1.37 (3H, d, J = 6.5, 2-Me), 1.32 (3H, d, J = 7.0, 4-Me), 1.3 (1H, m, H-14), 1.27 (3H, d, J = 6.5, 5'-Me), 1.25 (1H, m, 7-H), 1.12 (3H, d, J =6.0, 8-Me), 1.11 (3H, d, J = 6.5, 12-Me), 1.07 (3H, d, J = 6.0, 6-Me), 0.91 (3H, 1, J -7.2, 1 + Me); high resolution FAB MS calculated for $C_{28}H_{36}NO_2$ ($M+H^-$) 540.3172.found 540.3203.

One aspect of the invention relates to an isolated DNA strand containing the *cal*G gene and having the DNA sequence SEQ ID. NO.: 5. Another aspect of the invention is the protein, CalG, having amino acid sequence SEQ ID. No.: 6. Based on BLAST analysis it was presumed that *cal*G encoded a 4,6-dehydratase. Dehydratases had been characterized from *E. coli, Salmonella* and *Streptomyces,* (Thompson, M. et al., *J. Gen. Microbiol.*, 138, 779-786 (1992); Vara, J.A., et al., *J. Biol. Chem., 263,* 14992-14995 (1988)), and analogous NDP-D-glucose 4,6-dehydratases had been characterized from a variety of organisms. Liu, H.-w., et al., *Ann. Rev. Microbiol., 48,* 223-256 (1994); Hallis, T.M., et al., *Acc. Chem. Res.*, in press (1999). Based upon these prior studies, it was known that the overall transformation catalyzed by 4,6-dehydratases is an intramolecular oxidation-reduction where an enzyme-bound NAD^- receives the 4-H as a hydride in the oxidative half-reaction and passes the reducing equivalents to C-6 of the dehydration

product in the reductive half-reaction. Thus, it appears that Cal G is necessary for the formation of the aryltetrasaccharide 4,6-dideoxy-4-hydroxylamino-D-glucose moiety. CalG appears to be a TDP-D-glucose 4,6-dehydratase which catalyzes the conversion of intermediate 13 into intermediate 30. (See figure 5). Another aspect of the invention is an expression vector containing *cal*G or a fragment of *cal*G encoding for a bioactive molecule. There is also provided a transformed host cell, preferably bacteria, more preferably, *E. coli*, containing *cal*G or a fragment of *cal*G encoding for a bioactive molecule.

There is also disclosed an isolated DNA strand containing the *cal*S gene. Based on sequence homology with other P450-oxidases, CalS appears to be a P450-oxidase homolog which performs the oxidation of intermediate 39 to intermediate 42 (figure 5). The oxidation may occur at the nucleotide sugar level or hydroxylamine formation after the sugar has been transferred to the aglycone. There is also provided an expression vector containing the *cal*S gene or a fragment of *cal*S encoding for a bioactive molecule. There is also provided a transformed host cell, preferably bacteria, more preferably, *E. coli*, containing *cal*G or a fragment of *cal*G encoding for a bioactive molecule.

The present invention allows genetic manipulation of the biosynthetic gene cluster to produce calicheamicin analogs. The present invention provides for producing calicheamicin analogs by constructing deletions or substitutions of the genes involved in biosynthesis of the aryltetrasaccharide. The invention further provides for *in vitro* glycosylation by altering the glycosylation pattern of calicheamicin (via a glycosyltransferase) to produce additional analogs. The invention also provides for alteration of the calicheamicin aglycone by genetic manipulation of the genes encoding the

biosynthesis of the warhead. Genetic manipulation, such as producing deletions or substitutions are performed using methods known in the art.

The invention provides for a method of purifying calicheamicin through affinity chromatography. *CalC*, because of its homology with calicheamicin functions as a calicheamicin-sequestering/binding protein. Affinity chromatography is performed using methods known in the art.

The invention relates to the expression of the genes located in the biosynthetic gene cluster by using methods known in the art to insert the genes into a suitable expression vector and operably linking the gene to regulatory sequences to control expression of the gene to produce the protein encoded by the inserted gene. The present invention also provides for expression of biologically active proteins by inserting fragments of genes selected from the biosynthetic gene cluster, which encode for biologically active proteins, into a suitable expression vector, using methods known in the art. The genes would be operably linked to regulatory sequences to control their expression.

EXAMPLES

Example 1

To rapidly elucidate the nucleotide sequence, thermocycle sequencing was accomplished from pUC- or pBluescript-based subclones (using M13 primers and primer walking) as well as directly from isolated cosmids (via primer walking). Nucleotide sequence data was acquired using two Applied Biosystems automated 310 genetic analyzers and sequences were subsequently assembled using the Applied Biosystems AutoAssembler™ DNA sequence assembly software. Dear, S., et al., *Nucl Acids Res., 14,*

3907-3911 (1991); Huang, X., *Genomics, 14*, 18-25 (1992). Orf assignments were accomplished using a combination of the computational programs MacVector™ 6.0 and Brujene. MacVector is a commercially available software package which provides the ability to construct a *Micromonospora* codon bias table (from known *Micromonospora* sequences) and subsequently use this codon bias table to search for optimal orfs. Fickett, J.W., *Nucleic Acids Research, 10*, 5303-5318 (1982). Alternatively, the shareware program Brujene was specifically designed for streptomycetes and assigns priority to orfs that illustrate a consistency high G/C% in the wobble position.

Example 2: Isolating and Characterizing *cal*C

To isolate the gene(s) responsible for calicheamicin resistance in *Micromonospora*, clones conferring calicheamicin resistance were selected by growth of a *Micromonospora* genomic bifunctional cosmid library on LB plates containing ampicillin (50 $\mu g\ ml^{-1}$) and calicheamicin (0.25 $\mu g\ ml^{-1}$). In this selection, six clones (3a, 4a, 4b, 10a, 13a and 16a) displayed resistance to calicheamicin. Restriction mapping of these clones localized the desired phenotype to a ~2kb *Pst*I-*Sac*I fragment of DNA. (Figure 2). Maximum tolerated concentrations of calicheamicin on the LB plates was ascertained. The results are as follows:

Cosmid or Plasmid	**Maximum tolerated concentration of calicheamicin**
cosmids 3a, 4a, 10a, 13a, and 16a	0.5 $\mu g\ ml^{-1}$
pJT1214 and pJT1232	5.0 $\mu g\ ml^{-1}$
pRE7	20.0 $\mu g\ ml^{-1}$

induced pRE7	50.0 μg ml^{-1}
pJT1224 , pAP6, Pre1, and control plasmids pUC18, pBluescript, and pMAL-C2	<0.01 μg ml^{-1}

Nucleotide sequence analysis of the *Pst*I-*Sac*I fragment suggested that it contained four possible orfs. The proximal 1 kb of this fragment carried a single orf (*cal*D). The distal 1 kb presented three overlapping orf candidates (*cal*C, *cal*C', and *cal*C"). Computer translation of these three orfs (*cal*C, *cal*C', and *cal*C") was performed and subsequent BLAST analysis of their corresponding proteins, CalC, CalC', and CalC", respectively, revealed no homology with known proteins, while the translation of gene *cal*C displayed a weak alignment with apoproteins of the chromoprotein enediynes. Translation of *cal*D to its respective protein, CalD, revealed the presence of three amino acid motifs typically conserved in S-adenosylmethionein-utilizing *O*-methyltransferases. Therefore it was hypothesized that *cal*D was not responsible for calicheamicin resistance. To rule out *cal*D as being responsible for calicheamicin resistance, a subclone was engineered (pjT1224) to contain an intact *cal*D, but truncated *cal*C, *cal*C', and *cal*C" genes. This subclone was not able to confer resistance to calicheamicin. Next, a subclone containing the *cal*C region was constructed (pjT1232). This clone conferred calicheamicin resistance. See above chart. Subclones containing *cal*C' (pAP6) and *cal*C" (pRE1) were constructed and tested for calicheamicin resistance. These clones could not confer resistance to calicheamicin. See above chart.

To ascertain the amino acid sequence of CalC and learn its properties, *calC* was cloned into a pMAL-C2 vector. (pMAL-C2 by itself could not confer calicheamicin resistance. See above chart.) The resulting plasmid, pRE7, which contained *calc*, conferred resistance to calicheamicin. See above chart. Plasmid pRE7 was then induced with Isopropyl Beta-D-thiogalactoside ("IPTG") to overexpress CalC. Induced pRE7 conferred resistance to calicheamicin and produced a maltose-binding protein CalC fusion protein (mbp-CalC). This resulting overexpression of CalC increased calicheamicin resistance 10^2-fold *in vivo*. See above chart.

Example 3: Expression of protein CalC

The protein mbp-CalC was overexpressed and purified for further analysis. The mbp-CalC was purified from pRE7/*E. coli* to homogeneity as judged by SDS-PAGE. An overnight LB culture (containing 50 mg ml^{-1} ampicillin and 50 ng ml^{-1} calicheamicin from a fresh pRE7/*E. coli* colony was grown at 37 °C, 250 rpm to an A_{600}=0.5, induced with 0.5 mM IPTG and growth continued overnight. Cells were harvested (4,000 x g, 4 °C, 20 minutes), resuspended in buffer A (50mM Tris-Cl, pH 7.5, 200 mM NaCl, 1mM EDTA) and disrupted by sonication. The cell debris was removed by centrifugation (5,000xg, 4°C, 20 minutes). The supernatant was applied to an amylose affinity column (1.5 x 7.0 cm, 1 mL min^{-1}). The desired mbp-CalC protein was eluted with buffer A containing 10 mM maltose. The eluate was concentrated and chromatographed on an S-300 column (50mM Tris-Cl, pH 7.5, 200 mM NaCl). Active fractions were used immediately or frozen at -80°C for storage.

Example 4: Analysis of Protein CalC

The purified mbp-CalC was then analyzed for metal content. Purified mbp-CalC displayed a yellow color in concentrated form and subsequent metal analysis using inductively coupled plasma atomic mass spectrometry ("ICP-MS"), revealed the presence of iron (Fe). Determination of the Fe stoichiometry, accomplished in conjunction with quantitative amino acid hydrolysis, indicated 2.23 ± 0.3 mol Fe per mol mbp-CalC (based upon the monomeric molecular weight of 63,576 dalton calculated from the known nucleotide sequence of the mbp-*calC* gene fusion, which is consistent with the determined subunit molecular weight determined by SDS-PAGE). The precise mbp-CalC concentration was determined by quantitative amino acid hydrolysis by the Rockefeller University Protein/DNA Technology Center. Trace metal content of an aliquot of the hydrolysate was subsequently determined by ICP-MS on four distinct mbp-CalC preparations with buffer alone and/or maltose-binding protein alone analyzed in parallel as controls. These results were independently confirmed by methodologies used for spectrophotometric iron determination. Fish, W.W.; *Meth. Enzymol.* 1988, 158, 357-364. The electronic absorption spectrum of mbp-CalC is shown in Figure 4. In addition, to the A_{280} protein absorbance (ϵ_{280} = 99,300 M^{-1} cm^{-1}), a clear absorbance maxima at 411 nm (ϵ_{411} = 6,000 M^{-1} cm^{-1}) can be observed. Electron para magnetic resonance ("EPR") was performed to ascertain the metal content of CalC. The X-band EPR measurements on the oxidized CalC proteins exposed a standard rhombic EPR signal at g = 4.3 (E/D = 0.33)(Figure 4, inset). The metal content was 90 ± 10 μM Fe (approximately 72 ± 10% of total iron as seen by ICP-MS, Figure 4. The spectroscopic evidence indicates the presence

of a mononuclear Fe^{+3} center in CalC is consistent with the lack of cysteins in the primary sequence of CalC. *See* Palmer, G., *Biochem. Soc. Trans.* 1985, 13, 548-560.

Example 5: Verification of CalC's calicheamicin resistance

Given that calicheamicin leads to double strand DNA cleavage and CalC provides calicheamicin-resistance *in vivo*, it was expected that the addition of CalC to an *in vitro* calicheamicin-induced DNA cleavage assay would inhibit DNA cleavage. To test this theory, preliminary assays were performed with supercoiled pBlusecript plasmid DNA ("pBS") as the template, and dithiothreitol ("DTT") as the reductive initiator. In a typical assay, purified mbp-CalC (15.0 nM) and 30.0 nM calicheamicin were preincubated for 15 min. in a total volume of 25 µl 40 mM Tris-Cl, pH 7.5, at 37 °C. Then 2.5 µl 10mM DTT stock solution was added to the assay solution, and the assay was incubated an additional 1 hour at 37°C. DNA fragmentation was assessed by electrophoresis on a 1% agarose gel stained with ethidium bromide. Using this assay, it was found that mbp-CalC could completely inhibit calicheamicin-induced DNA cleavage at concentrations nearing 10^3-fold excess of calicheamicin. Preincubation of mbp-CalC and DTT, protein removal via forced dialysis, and the subsequent use of the DTT solution as reductant did not noticeably affect the amount of DNA cleavage.

As indicated in Figure 4(b), no DNA cleavage was observed in the absence of DTT or calicheamicin (lanes a and b), while efficient cleavage was demonstrated in the presence of DTT and calicheamicin (lane c). As expected, the addition of mbp-CalC completely inhibited calicheamicin-induced DNA cleavage (lane f) while the addition of mbp alone (lane d) as a control, failed to inhibit calicheamicin-induced DNA cleavage. Furthermore,

preincubation of mbp-CalC with DTT (not shown), or *apo*-mbp-CalC (lacking the Fe cofactor)(lane e), also failed to inhibit calicheamicin-induced DNA cleavage. However, the addition of Fe^{+2} or Fe^{+3} to the *apo*-mbp-CalC assay could reconstitute CalC activity (lane g). Reconstitution of *apo*-mbp-CalC was accomplished by preincubation with 1 mM $FeSO_4$ (Fe^{+2}) or $FeCl_3$ (Fe^{+3}) prior to the activity assay as previously described.

Example 6: Production of methymycin/pikromycin-calicheamicin hybrid compounds

The 1.2 kb *calH* gene was amplified by polymerase chain reaction (PCR) from $pJST1192_{kpn7}$, which is a subclone containing a 7.0 kb *Kpn*I fragment of cosmid 13a. The amplified gene was cloned into the *Eco*RI/*Xba*I site of the expression vector pDHS617. This expression vector contains an apramycin resistance marker. The plasmid pDHS617 was derived from pOJ1446 (Bierman, M. et al., *Gene* 1992, 116, 43-49). A promoter sequence from the *S. venezuela* methymycin/pikromycin cluster was incorporated in the plasmid to drive the expression of foreign genes in *S. venezuela*. The resulting plasmid, pLZ-C242 (containing the *calH* gene insert and the promoter sequence) was introduced by conjugal transfer using *E.coli* S 17-1 into a previously constructed *S. venezuela* mutant, desI. (Borisova, S. et al., *Org. Lett.* 1999, 1, 133-136). In the DesI mutant, the *des*I was replaced by the ncomycin resistance gene, which confers resistance to kanamycin The PLS-C242-containing *S. venezuela*-DesI colonies were identified on the basis of their resistance to apramycin antibiotic. One of these positive colonies, DesI/calH-1 was grown in 100 ml of seed medium at 29°C for 48 hours and then inoculated and grown in five Liters of vegetative medium. Cane, D.E., et al., *J. Am. Chem. Soc.*, 1993, 115, 522-526. The culture was centrifuged to remove cellular debris and mycella. The supernatant was

adjusted to pH 9.5 with concentrated KOH, followed by chloroform extraction. The crude products (700 mg) were subjected to flash chromatography on silica gel using a gradient of 1-20% methanol in chloroform. A major product, 10-deoxymethynolide (ca. 400 mg), and a mixture of two minor macrolide compounds were obtained. The two macrolides were further purified by HPLC on a C_{18} column using an isocratic mobile phase of acetonitrile/H_2O (1:1). They were later identified as compound (11) and compound (12)(figure 7) by spectral anaylses.

Sequence listing 1 --calC gene:

ATGACTCAGGAGAAGACCGCACCGGCCGCGAAGAGCACGACCACCAAGAGCA
CCGCCGCGAAGAAGCCGAAGCCCCCGAACTACGACCCGTTCGTCCGGCACAG
CGTCACTGTCAAGGCCGACCGCAAGACCGCCTTCAAGACGTTCCTCGAAGGCT
TTCCGGAGTGGTGGCCGAACAACTTCCGCACCACCAAGGTCGGGGCCCCGCTG
GGCGTCGACAAGAAGGGCGGCCGCTGGTACGAGATCGACGAGCAGGGCGAGG
AGCACACCTTCGGCCTGATCCGGAAGGTGGACGAGCCGGACACGCTGGTCATC
GGCTGGCGGCTCAACGGCTTCGGCCGGATCGACCCGGACAACTCGAGCGAGTT
CACCGTGACCTTCGTGGCCGACGGCCAGAAGAAGACCCGGGTGGACGTCGAG
CACACCCACTTCGACCGGATGGGCACCAAGCACGCCAAGCGGGTCCGCAACG
GCATGGACAAGGGCTGGCCGACGATCCTCCAGTCGTTCCAGGACAAGATCGAC
GAGGAAGGGGCGAAGAAGTGA

Sequence Listing 2 --CalC protein:

(Note that in protein sequences amino acids are designated in one-letter code)

MTQEKTAPAAKSTTTKSTAAKKPKPPNYDPFVRHSVTVKADRKTAFKTFLEGFPE

WWPNNFRTTKVGAPLGVDKKGGRWYEIDEQGEEHTFGLIRKVDEPDTLVIGWRL

NGFGRIDPDNSSEFTVTFVADGQKKTRVDVEHTHFDRMGTKHAKRVRNGMDKG

WPTILQSFQDKIDEEGAKK

Sequence Listing 3 -- calH gene:

GTGGCAACTAGCGAGAGGGGTGTCATGATCCCGCTGTCCAAGGTCGCCATGTC
TCCGGACGTCAGCACCCGCGTCTCCGCCGTCCTGAGCAGTGGCCGGCTGGAGC
ACGGGCCGACCGTCGCCGAGTACGAGGCGGCCGTGGGCAGTCGTATCGGCAA
CCCCCGGGTGGTCTCGGTCAACTGCGGCACGGCCGGGCTCCACCTGGCGCTGA
GCCTCGCCGCGCGGCCGGGGGCCGGCGAGTCGGAGCACGACGGCCCGGGCGA
GGTGCTCACCACGCCGCTGACCTTCGAGGGCACGAACTGGCCGATCCTCGCCA
ACGGGCTGCGCATCCGGTGGGTGGACGTCGACCCGGCCACCCTCAACATGGAC
CTCGACGACCTGGCCGCGAAGATCTCGCCCGCCACCCGGGCCATCGTGGTGGT
CCACTGGCTCGGCTACCCGGTGGACCTCAACCGGCTGCGCGCCGTCGTGGACC
GGGCCACGGCGGGATACGACCGCCGCCCGCTGGTCGTGGAGGACTGCGCGCA
GGCGTGGGGCGCCACCTACCGGGGCGCGCCGCTGGGCACGCACGGCAACGTC
TGCGTGTACAGCACCGGCGCGATCAAGATCCTGACGACCGGCAGCGGCGGCTT
CGTCGTGCTGCCCGACGACGACCTGTACGACCGGCTCCGGCTGCGCCGCTGGC
TCGGCATCGAGCGGGCGTCGGACCGGATCACCGGCGACTACGACGTCGCCGA
GTGGGGCTACCGGTTCATCCTCAACGAGATCGGCGGGGCGATCGGCCTGTCCA
ACCTGGAACGCGTCGACGAGCTGCTGCGCCGGCACCGGGAGAACGCCGCGTT
CTACGACAAGGAACTGGCCGGCATCGACGGCGTCGAGCAGACCGAGCGGGCC
GACGACCGGGAGCCCGCGTTCTGGATGTACCCGCTGAAGGTCCGCGACCGTCC
CGCCTTCATGCGCCGGCTGCTCGACGCCGGCATCGCCACCAGCGTCGTGTCGC
GCCGCAACGACGCGCACAGCTGCGTCGCGTCGGCCCGCACCACCCTGCCCGGG
CTGGACCGGGTGGCGGACCGCGTGGTCCACATCCCGGTGGGCTGGTGGCTCAC
CGAGGACGACCGCTCCCACGTCGTCGAAACGATCAAGTCCGGCTGGTGA

Sequence Listing 4 --CalH protein:

MATSERGVMIPLSKVAMSPDVSTRVSAVLSSGRLEHGPTVAEYEAAVGSRIGNPR
VVSVNCGTAGLHLALSLAARPGAGESEHDGPGEVLTTPLTFEGTNWPILANGLRIR
WVDVDPATLNMDLDDLAAKISPATRAIVVVHWLGYPVDLNRLRAVVDRATAGY
DRRPLVVEDCAQAWGATYRGAPLGTHGNVCVYSTGAIKILTTGSGGFVVLPDDD
LYDRLRLRRWLGIERASDRITGDYDVAEWGYRFILNEIGGAIGLSNLERVDELLRR
HRENAAFYDKELAGIDGVEQTERADDREPAFWMYPLKVRDRPAFMRRLLDAGIA
TSVVSRRNDAHSCVASARTTLPGLDRVADRVVHIPVGWWLTEDDRSHVVETIKS
GW

Sequence Listing 5 --calG gene:

GTGCCCAGATCCCTGGTCACCGGCGGCTTCGGCTTCGTCGGCAGTCACGTCGT

CGAACGGCTGGTCCGCCGGGGTGACGAGGTCGTCGTCTACGACCTCGCCGACC

CGCCGCCCGACCTGGAGCACCCGCCGGGCGCGATCCGGCACGTCCGCGGCGA

CGTCCGGGACGCCGACGGGCTGGCGGCCGCCGCCACCGGCGTGGACGAGGTC

TACCACCTCGCGGCGGTCGTCGGCGTCGACCGGTACCTCAGCCGGCCGCTGGA

CGTGGTCGAGATCAACGTGGACGGCACCCGGAACGCGTTGCGCGCCGCACTG

CGCGCCGGTGCCCGGGTCGTGGTGTCCAGCACCAGCGAGGTGTACGGGCGCA

ATCCGCGGGTGCCGTGGCGGGAGGACGACGACCGGGTGCTCGGCAGCACGGC

GACGGACCGGTGGTCGTACTCGACGAGCAAGGCGGCGGCCGAGCACCTGGCC

TTCGCCTTCCACCGGCAGGAGGGCCTGCCGGTGACGGTGCTGCGGTACTTCAA

CGTCTACGGCCCACGCCAGCGCCCGGCGTACGTCCTCAGCCGCACCGTCGCCC

GCCTGCTGCGGGGCGTTCCGCCCGTGGTGTACGACGACGGCCGCCAGACGCGG

TGCTTCACCTGGATCGACGAGGCGGCCGAGGCGACCCTGCTGGCCGCCGCCCA

CCCGCGGGCCGTCGGCGAGTGTTTCAACATCGGCAGCAGCGTGGAGACCACC

GTCGCCGAGGCGGTCCGGCTGGCCGGCACGGTGGCCGGGGTGCCGGTGGCGG

CCCAGACCGCGGACACCGGAGCCGGGCTCGGCGCCCGCTACCAGGACATTCC

CCGCCGCGTACCGGACTGCGGCAAGGCCGCCGCGCTGCTGGACTGGCGGGCC

CGGGTGCCGCTGGTGACCGGCCTGCGCCGGACCGTCGAGTGGGCCCGCCGCA

ACCCGTGGTGGACCGCCCAGGCCGACGACGGACTGGTCGTCAGGTAG

Sequence Listing 6 -- CalG protein:

MPRSLVTGGFGFVGSHVVERLVRRGDEVVVYDLADPPPDLEHPPGAIRHVRGDV
RDADGLAAAATGVDEVYHLAAVVGVDRYLSRPLDVVEINVDGTRNALRAALRA
GARVVVSSTSEVYGRNPRVPWREDDDRVLGSTATDRWSYSTSKAAAEHLAFAFH
RQEGLPVTVLRYFNVYGPRQRPAYVLSRTVARLLRGVPPVVYDDGRQTRCFTWI
DEAAEATLLAAAHPRAVGECFNIGSSVETTVAEAVRLAGTVAGVPVAAQTADTG
AGLGARYQDIPRRVPDCGKAAALLDWRARVPLVTGLRRTVEWARRNPWWTAQ
ADDGLVVR

CLAIMS

1. An isolated nucleic acid molecule from *Micromonospora echinospora* comprising a nucleic acid sequence encoding for a gene from a nonchromoprotein enediyne biosynthetic gene cluster, the protein coding region of said gene or a biologically active fragment of said gene.

2. The isolated nucleic acid molecule of Claim 1, wherein said gene is *cal*A, *cal*B, *cal*C, *cal*D, *cal*E, *cal*F, *cal*G, *cal*H, *cal*I, *cal*J, *cal*K, *cal*L, *cal*M, *cal*N, *cal*O, *cal*P, *cal*Q, *cal*R, *cal*S, *cal*T, *orf*1, *orf*2, *orf*3, *orf*4, *orf*5, *orf*6, *orf*7, or an IS-element gene.

3. The isolated nucleic acid molecule of Claim 1, wherein said molecule encodes two or more of said genes.

4. The isolated nucleic acid molecule of Claim 1, wherein said molecule encodes the full biosynthetic gene cluster.

5. The isolated nucleic acid molecule of Claim 1, wherein said nonchromoprotein enediyne is calicheamicin.

6. An isolated nucleic acid molecule capable of hybridizing with a nucleic acid from *Micromonospora echinospora* spp. *calichensis* encoding for one or more genes from a nonchromoprotein enediyne biosynthetic gene cluster.

7. The isolated nucleic acid molecule of Claim 6, wherein said molecule encodes a protein having the activity of at least one gene from the biosynthetic gene cluster.

8. The isolated nucleic acid molecule of Claim 6, wherein said gene is *cal*A, *cal*B, *cal*C, *cal*D, *cal*E, *cal*F, *cal*G, *cal*H, *cal*I, *cal*J, *cal*K, *cal*L, *cal*M, *cal*N, *cal*O, *cal*P,

calQ, *calR*, *calS*, *calT*, *orf1*, *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, or an IS-element gene.

9. The isolated nucleic acid molecule of Claim 1, comprising SEQ ID No.1.

10. The isolated nucleic acid molecule of Claim 1, comprising SEQ ID No.3.

11. The isolated nucleic acid molecule of Claim 1, comprising SEQ ID No.5.

12. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a P_{450} oxidase from *Micromonospora echinospora* spp. *calichensis*.

13. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a membrane transporter from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

14. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for an *O*-methyltransferase from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

15. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a glycosyltransferase from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

16. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a *N,N*-dimethyltransferase from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

17. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a dipeptide transporter from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

18. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a L-cysteine/cystine C-S-lyase from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

19. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for an oligopeptide transporter protein from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

20. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a regulatory protein from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

21. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a hexopyranosyl-2-3-reductase from *Micromonospora echinospora* spp. *calichensis*.

22. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a desaturase from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

23. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for an UDP-D-glucose 6-dehydrogenase from *Micromonospora echinospora* spp. *calichensis*.

24. The isolated nucleic acid molecule of Claim 1, which encodes a polypeptide encoding for a transcriptional regulator from a gene cluster of *Micromonospora echinospora* spp. *calichensis* coding for calicheamicin biosynthesis.

25. An expression vector comprising a nucleic acid molecule encoding a protein coding sequence, wherein the nucleic acid molecule is selected from any of Claims 1 through 24.

26. The expression vector of Claim 25, wherein said nucleic acid molecule is operably linked to regulatory sequences to control expression of said protein.

27. The expression vector of Claim 26, wherein the regulatory sequence is a *Streptomyces* promoter.

28. A host cell transformed with a nucleic acid molecule from any one of Claims 1 through 24.

29. A host cell transformed with a nucleic acid molecule from Claim 25.

30. A host cell transformed with a nucleic acid molecule from Claim 26.

31. The host cell of Claim 28, wherein said host cell is a bacterium, yeast, insect, plant, fungi, or mammalian cell.

32. The host cell of Claim 28, wherein the host bacteria is *E. coli* or *Streptomyces*.

33. A cosmid comprising an isolated nucleic acid molecule from *Micromonospora echinospora* spp. *calichensis*, comprising a nucleic acid sequence encoding for a nonchromoprotein enediyne biosynthetic gene cluster.

34. The cosmid of Claim 33, wherein said sequence encodes *cal*A, *cal*B, *cal*C, *cal*D, *cal*E, *cal*F, *cal*G, *cal*H, *cal*I, *cal*J, *cal*K, *cal*L, *cal*M, *cal*N, *cal*O, *cal*P, *cal*Q, *cal*R, *cal*S, *cal*T, *orf*1, *orf*2, *orf*3, *orf*4, *orf*5, *orf*6, *orf*7, and an IS-element gene.

35. A method of expressing a protein comprising culturing a host cell with an expression vector from Claim 25 for and incubating under time and conditions that allow for protein expression.

36. The method of Claim 35, wherein said host cell is a bacterium, yeast, insect, plant, fungi, or mammalian cell.

37. A method of purifying calicheamicin using affinity chromatography, comprising providing a solution containing calicheamicin to an affinity column having CalC bound thereto, and recovering calicheamicin.

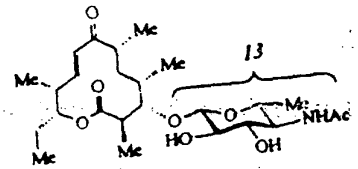
38. A polypeptide comprising an amino acid sequence SEQ ID. No.: 2.

39. A polypeptide comprising an amino acid sequence SEQ ID. No.: 4.

40. A polypeptide comprising an amino acid sequence SEQ ID. No.: 6.

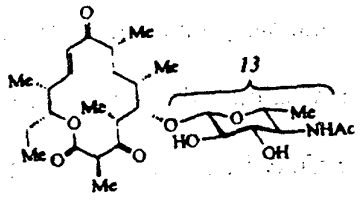
41. A method of conferring calicheamicin resistance to a subject comprising obtaining cells from the subject, transforming the cells with the calicheamicin self resistance gene and returning the cells to the subject.

42. A compound having the structure:



11

43. A compound having the structure:



12

Summary of cosmid clones isolated from *M. echinospora* genomic library.

clone[a]	type I PKS genes[b]	type II PKS genes[b]	deoxy sugar genes[b]	resistance ($\mu g\ mL^{-1}$)[c]
3a	N.D.[d]	N.D.[d]	N.D.[d]	0.5
4a	N.D.[d]	N.D.[d]	N.D.[d]	0.5
4b	+	+	+	0.5
10a	+	+	+	0.5
13a	+	+	+	0.5
16a	N.D.[d]	N.D.[d]	N.D.[d]	0.5
56	+	+	+	0.1
58	-	-	+	< 0.01
60	+	+	+	0.05
66	-	-	+	0.04
puc18/pBluescript[e]	-	-	-	< 0.01

Fig. 1

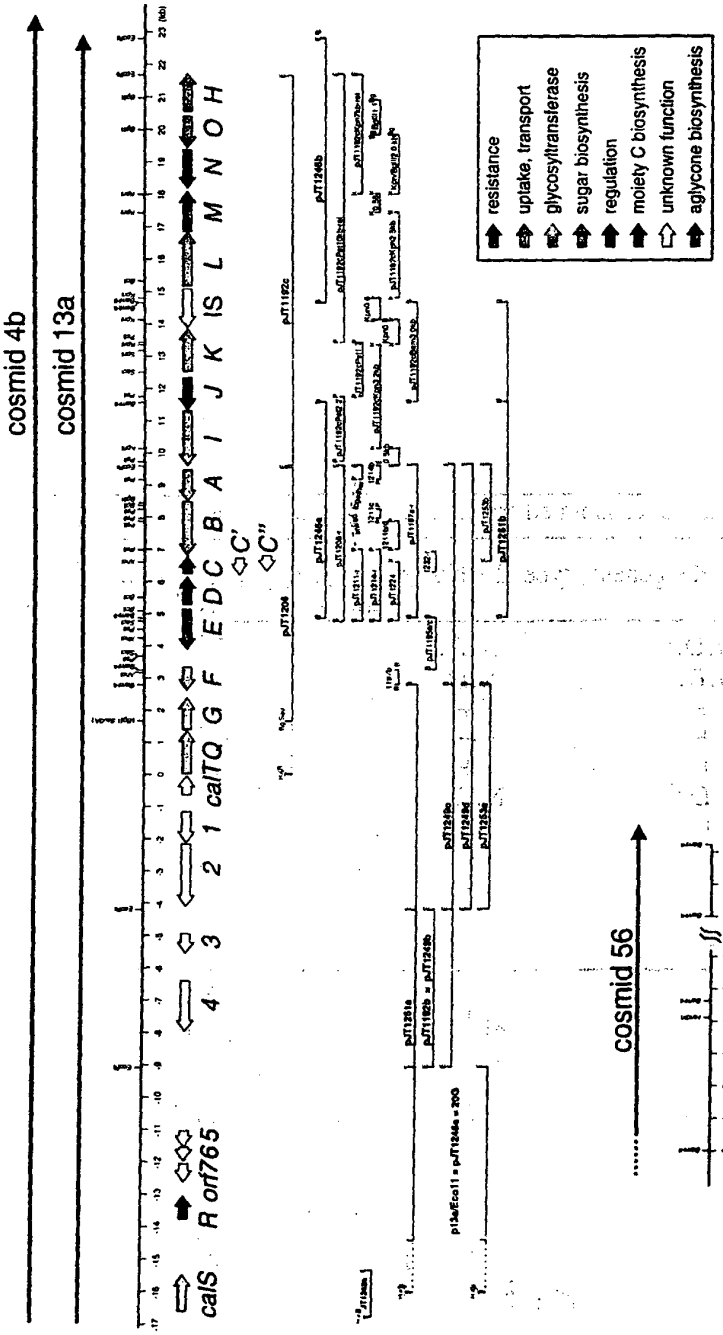


Fig. 2

putative polypeptide	number of amino acids	proposed function or sequence similarity detected[a]	probability[b]	start/stop codons	best match
CalA	328	membrane transporter (ATP-binding)	$5.4x10^{-123}$	ATG/TGA	DrrA[97]
CalB	561	membrane transporter	$5.5x10^{-70}$	ATG/TGA	DrrB[97]
CalC	181	calicheamicin resistance protein	confirmed[c]	ATG/TGA	*section C.5.2.*
CalD	263	*O*-methyltransferase	$1.1x10^{-99}$	ATG/TGA	AveBVII[98]
CalE	420	glycosyltransferase	$4.7x10^{-30}$	GTG/TAG	EryCII[99]
CalF	245	*N,N*-dimethyltransferase	$1.5x10^{-78}$	ATG/TGA	DesVI[100]
CalG	990	TDP-D-glucose 4,6-dehydratase	confirmed[c]	GTG/TAG	*section C.4.3.*
CalH	338	perosamine synthetase	confirmed[c]	GTG/TGA	*section C.4.4.*
CalI	568	dipeptide transporter	$1.7x10^{-24}$	GTG/TGA	DciAE[101]
CalJ	332	*O*-methyltransferase	$1.0x10^{-37}$	ATG/TGA	DmpM[102]
CalK	440	L-cysteine/lysine C-S-lyase	$1.6x10^{-28}$	GTG/TGA	C-DES[103]
CalL	562	oligopeptide transporter protein	$9.5x10^{-14}$	ATG/TGA	OppA[104]
CalM	416	regulatory protein		GTG/TGA	
CalN	398	glycosyltransferase	$3.4x10^{-79}$	ATG/TGA	Ole1[105]
CalO	331	hexopyranosyl-2,3-reductase	$4.9x10^{-139}$	ATG/TGA	EryBII[99]
CalP	(179)[d]	desaturase	$5.7x10^{-7}$	/TGA	CrtI[106]
CalQ	453	UDP-D-glucose 6-dehydrogenase	$6.7x10^{-74}$	GTG/TGA	gene 836[107]
CalR	265	regulatory protein	$3.7x10^{-13}$	ATG/TGA	KorSA[108]
CalS	1113	P_{450} oxidase	$2.9x10^{-66}$	GTG/TGA	BioI[109]
CalT	280	transcriptional regulator	$1.7x10^{-14}$	ATG/TGA	SC5C7.03[110]
orf1	322	unknown		ATG/TGA	
orf2	654	unknown		ATG/TGA	
orf3	209	unknown		ATG/TGA	
orf4	521	unknown		GTG/TAA	
orf5	175	unknown		ATG/TGA	
orf6	139	unknown		ATG/TGA	
orf7	187	unknown		GTG/TGA	
IS-element	1209 bp	insertional element	$5.7x10^{-108}$		IS*1136*[111]

Fig. 3

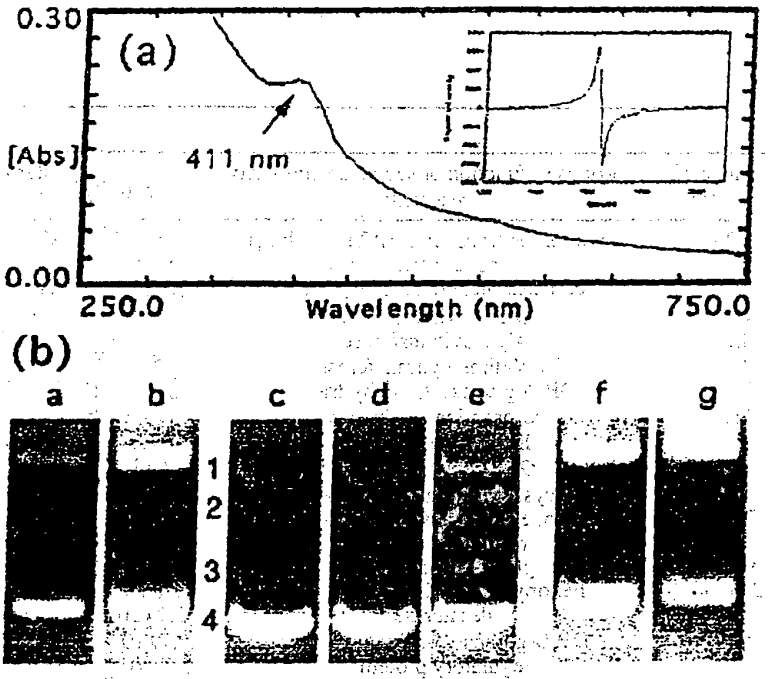


Fig. 4

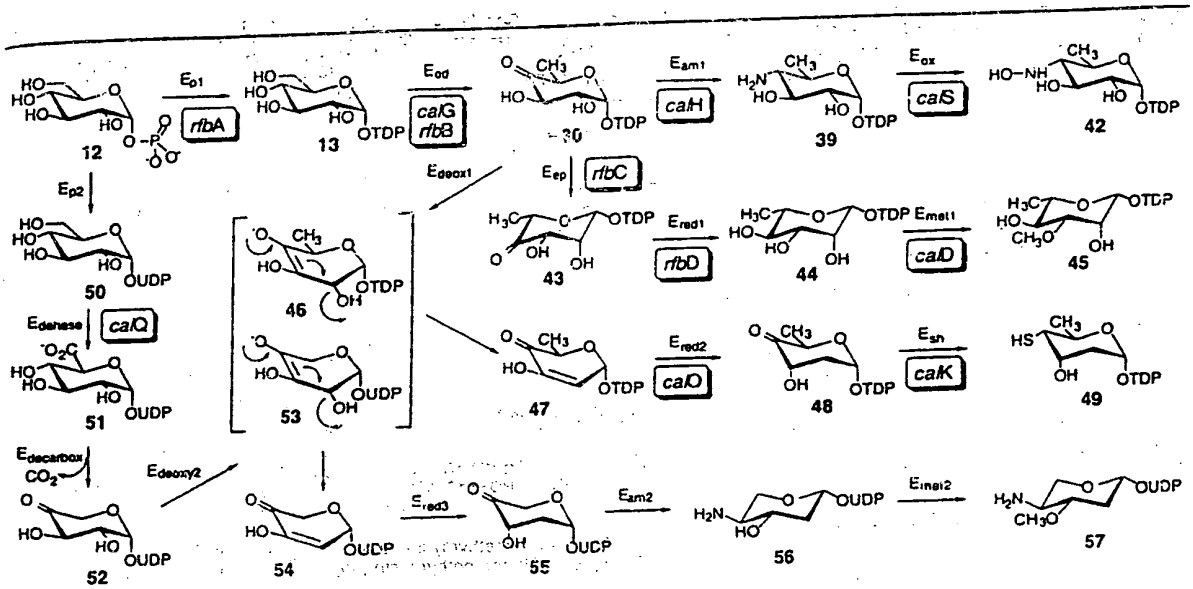


Fig. 5

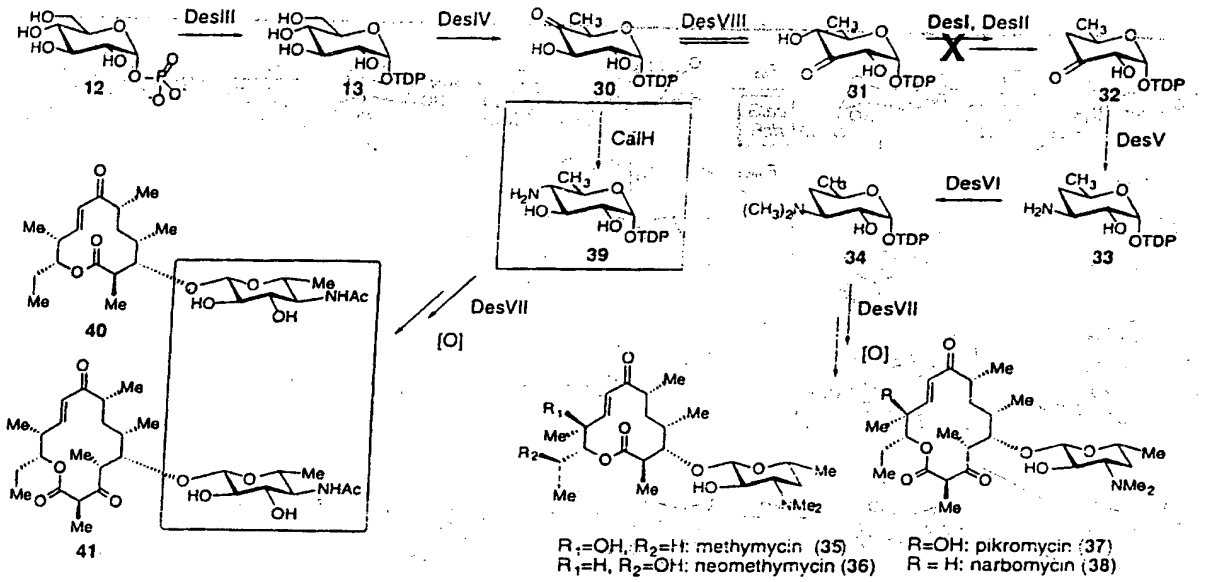


Fig. 6

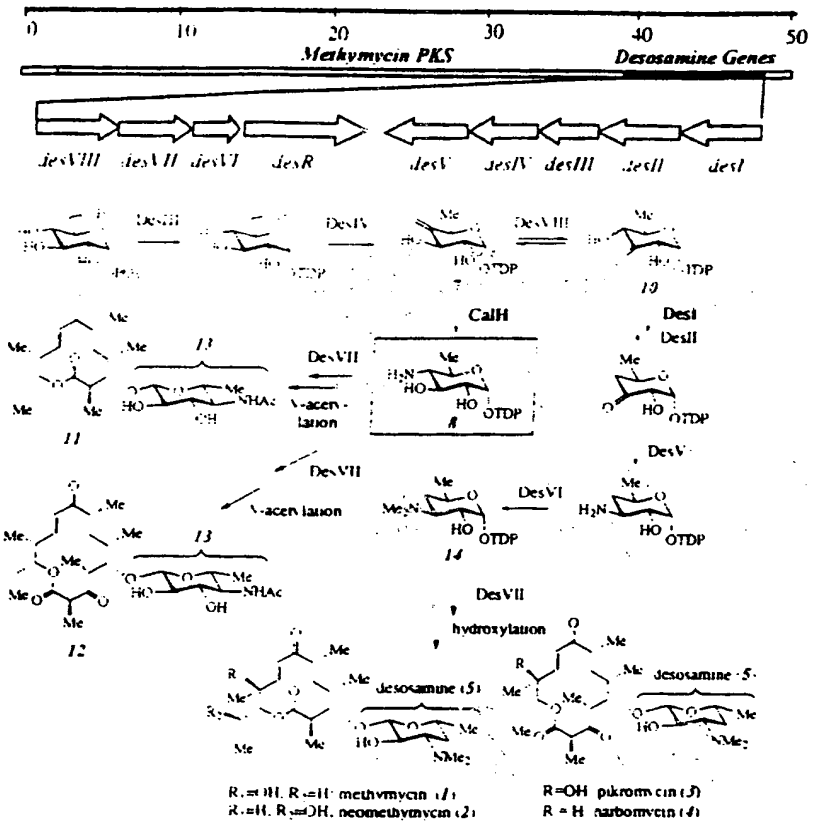


Fig. 7

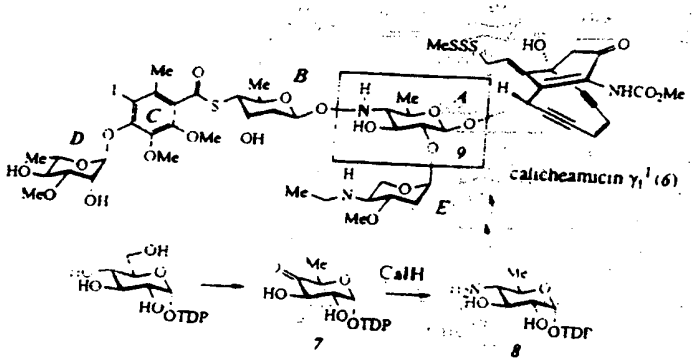


Fig. 8

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12P 19/60, C12N 9/00, 1/20, 15/00, C07H 21/04, A23J 1/00	A3	(11) International Publication Number: **WO 00/37608** (43) International Publication Date: 29 June 2000 (29.06.00)

(21) International Application Number: PCT/US99/29110

(22) International Filing Date: 7 December 1999 (07.12.99)

(30) Priority Data:
60/111,325 7 December 1998 (07.12.98) US

(71) Applicant: SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH [US/US]; 1275 York Avenue, New York, NY 10021 (US).

(72) Inventor: THORSON, Jon; 1161 York Avenue, Apartment 3E, New York, NY 10021 (US).

(74) Agents: DELUCIA, Richard, L. et al.; Kenyon & Kenyon, One Broadway, New York, NY 10004 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(88) Date of publication of the international search report:
23 November 2000 (23.11.00)

(54) Title: *MICROMONOSPORA ECHINOSPORA* GENES ENCODING FOR BIOSYNTHESIS OF CALICHEAMICIN AND SELF-RESISTANCE THERETO

(57) Abstract

An isolated gene cluster of *Micromonospora echinospora* which codes for calicheamicin biosynthesis. The biosynthetic gene cluster contains genes encoding for proteins and enzymes used in the biosynthetic production of calicheamicin, including the aryltetrasaccharide and aglycone. The gene cluster also includes the gene conferring calicheamicin resistance. The invention also provides isolated genes of the biosynthetic cluster and their corresponding proteins. In addition, the invention relates to DNA hybridizing with the calicheamicin gene cluster and the isolated genes of that cluster. Expression vectors containing genes of the biosynthetic gene and their functional variants are also provided. The invention also relates to host cells conjugated with DNA isolated from the *Micromonospora echinospora* spp. *calichensis* genome.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29110

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12P 19/60; C12N 9/00, 1/20, 15/00; C07H 21/04; A23J 1/00
US CL :435/76, 69.1, 183, 252.3, 320.1; 536/23.1, 23.2; 530/412
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/76, 69.1, 183, 252.3, 320.1; 536/23.1, 23.2; 530/412

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

West and STN. STN files included medline, uspatfull, caplus, biosis & scisearch. Search terms included micromonospora and calcheamicin gene, enediyne biosynthesis, etc.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,712,146 A (KHOSLA et al.) 27 January 1998, see the entire document.	1-8, 12-37, 41-43
A	US 5,276,159 A (SMITH et al.) 04 January 1994, see the entire document.	1-8, 12-37, 41-43

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
17 AUGUST 2000	07 SEP 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Auth rized officer TEKCHAND SAIDHA Telephone N . (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29110

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 9-11 & 38-40
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

Claims are to specific SEQ ID NOS. and no CRF is provided. Therefore the specific sequences cannot be searched.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ N protest accompanied the paym nt f additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29110

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-8 and 12-36, drawn to nucleic acid encoding enzymes of calicheamicin biosynthesis, vectors, host cells and a method of expressing or making the enzymes or proteins.
Group II, claim 37, drawn to a method of purifying calicheamicin.
Group III, claim 41, drawn to a method of conferring calicheamicin resistance.
Group IV, claim 42, drawn to the structure of a compound.
Group V, claim 43, drawn to the structure of a second compound.

The inventions listed as Groups I-V do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I has a special technical feature of nucleic acid encoding calicheamicin enzymes, and uses the nucleic acid for making the protein, which group II, IV-V do not share. Nucleic acid of Group I further does not share the special technical feature with group III which employs a nucleic acid for conferring calicheamicin self resistance, such a resistance is not conferred by the nucleic acid of Group I. Thus the various groups discussed above show a lack of unity of invention.